Robust multichannel sparse recovery

Esa Ollila

Department of Signal Processing and Acoustics
Aalto University, Finland

SUPELEC, Feb 4th, 2015

Aalto University

Menu

1. Introduction
2. Nonparametric sparse recovery
3. Simulation studies
4. Source localization with Sensor Arrays

Single Measurement Vector Model (SMV)

$$\boxed{\underset{M\times 1}{\mathbf{y}} = \underset{M\times N}{\mathbf{\Phi}}\ \underset{N\times 1}{\mathbf{x}} + \underset{N\times 1}{\mathbf{e}}}$$

- $\mathbf{\Phi} = \begin{pmatrix}\boldsymbol{\phi}_1 & \cdots & \boldsymbol{\phi}_N\end{pmatrix} = \begin{pmatrix}\boldsymbol{\phi}_{(1)} & \cdots & \boldsymbol{\phi}_{(M)}\end{pmatrix}^{\mathrm{H}}$,
  $M \times N$ known *measurement (design, system) matrix*
- $\mathbf{x} = (x_1, \ldots, x_N)^\top$, unobserved *signal vector*
- $\mathbf{e} = (e_1, \ldots, e_M)^\top$, random noise vector.

Objective

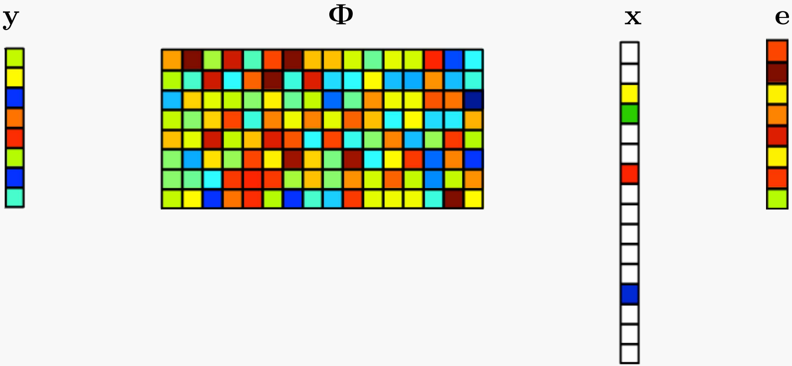
recover K-sparse or compressible signal from $\mathbf{y}$ knowing $\mathbf{\Phi}$, K

NOTE: typically, $M < N$ (ill-posed model) and $K \ll N$.

Single Measurement Vector Model (SMV)

$$\boxed{\underset{M \times 1}{\mathbf{y}} = \underset{M \times N}{\mathbf{\Phi}} \; \underset{N \times 1}{\mathbf{x}} + \underset{M \times 1}{\mathbf{e}}}$$

- If we can determine $\Gamma = \mathrm{supp}(\mathbf{x})$, then the problem reduces to conventional regression problem (more 'responses' than 'predictors'):



Multiple Measurement Vectors Model (MMV)

$$\mathbf{y}_i = \mathbf{\Phi}\mathbf{x}_i + \mathbf{e}_i, \quad i = 1, \ldots, Q$$

In matrix form

$$\boxed{\underset{M \times Q}{\mathbf{Y}} = \underset{M \times N}{\mathbf{\Phi}} \underset{N \times Q}{\mathbf{X}} + \underset{M \times 1}{\mathbf{E}}}$$

- $\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 & \cdots & \mathbf{y}_Q \end{pmatrix}$, observed measurement matrix
- $\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_Q \end{pmatrix}$, unobserved signal matrix
- $\mathbf{E} = \begin{pmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_Q \end{pmatrix}$, noise matrix

Objective

recover K-sparse (or compressible) signal matrix $\mathbf{X}$ knowing $\mathbf{Y}$, $\mathbf{\Phi}$, K.

$N \times Q$ matrix $\mathbf{X}$ is called K-sparse if $\|\mathbf{X}\|_0 = |\mathrm{supp}(\mathbf{X})| \leq K$, where

$$\Gamma = \mathrm{supp}(\mathbf{X}) = \bigcup_{i=1}^{Q} \mathrm{supp}(\mathbf{x}_i) = \{j \in \{1, \ldots, N\} : x_{jk} \neq 0 \text{ for some } k\}$$

Multiple Measurement Vectors Model (MMV)

Key idea:
signals are all predominantly supported on a common support set

$\Longrightarrow$ Joint estimation can lead both to computational advantages and increased reconstruction accuracy

See [Tropp, 2006, Chen and Huo, 2006, Eldar and Rauhut, 2010, Duarte and Eldar, 2011, Blanchard et al., 2014].

Applications

- EEG/MEG [Gorodnitsky et al., 1995]
- equalization of sparse communications channels [Cotter and Rao, 2002]
- blind source separation [Gribonval and Zibulevsky, 2010]
- source localization using sensor arrays [Malioutov et al., 2005]
- ...

Matrix norms

The mixed $\ell_{p,q}$ norm of $\mathbf{X}$ for $p, q \in [1, \infty)$

$$\|\mathbf{X}\|_{p,q} = \left(\sum_i \left(\sum_j |x_{ij}|^p \right)^{q/p} \right)^{1/q} = \left(\sum_i \|\mathbf{x}_{(i)}\|_p^q \right)^{1/q}.$$

- If $p = q \implies$ matrix p-norm $\|\mathbf{X}\|_{p,p} = \|\mathbf{X}\|_p$.
- ℓ_2 (Frobenius) norm $\|\cdot\|_2$ is denoted shortly as $\|\cdot\|$.

The mixed ℓ_∞-norms of $\mathbf{X} \in \mathbb{C}^{N \times Q}$

- $\|\mathbf{X}\|_{p,\infty} = \max_{i \in \{1,\ldots,N\}} \|\mathbf{x}_{(i)}\|_p$
- $\|\mathbf{X}\|_{\infty,q} = \left(\sum_i (\max_{j \in \{1,\ldots,Q\}} |x_{ij}|)^q \right)^{1/q}$.

The row ℓ_0 quasi-norm of $\mathbf{X}$

$$\|\mathbf{X}\|_0 = |\operatorname{supp}(\mathbf{X})| = \text{nr. of number of nonzero rows of } \mathbf{X}$$

Optimization problem

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{\Phi}\mathbf{X}\|^2 \quad \text{subject to} \quad \|\mathbf{X}\|_0 \leq K$$

is combinatorial (NP-hard). More feasible approaches:

- Optimization (Convex relaxation)
- Greedy pursuit:
  - Orthogonal Matching Pursuit (OMP)
  - Normalized Iterative Hard Thresholding (NIHT)
  - Normalized Hard Thresholding Pursuit (NHTP)
  - Compressive Sampling MP (CoSaMP)
- Guaranteed to perform very well under suitable conditions (restricted isometry property on $\mathbf{\Phi}$, non impulsive noise $\mathbf{e}$)
- NIHT is simple and scalable, convenient for problems where the support is of primary interest

Optimization problem

$$\boxed{\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{\Phi X}\|^2 \quad \text{subject to} \quad \|\mathbf{X}\|_0 \leq K}$$

is combinatorial (NP-hard). More feasible approaches:

- Optimization (Convex relaxation)
- Greedy pursuit:
  - Orthogonal Matching Pursuit (OMP)
  - Normalized Iterative Hard Thresholding (NIHT)
  - Normalized Hard Thresholding Pursuit (NHTP)
  - Compressive Sampling MP (CoSaMP)
- Guaranteed to perform very well under suitable conditions (restricted isometry property on $\mathbf{\Phi}$, non impulsive noise $\mathbf{e}$)
- NIHT is simple and scalable, convenient for problems where the support is of primary interest

Convex relaxation

IDEA: replace $\|\mathbf{X}\|_0$ by $\ell_{p,1}$ norm

- [Tropp, 2006] uses $p = \infty$:

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{\Phi X}\|^2 + \lambda \|\mathbf{X}\|_{\infty,1}$$

  where $\|\mathbf{X}\|_{\infty,1} = \sum_i \|\|\mathbf{x}_{(i)}\|_{\infty}$.

- [Malioutov et al., 2005] uses $p = 2$:

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{\Phi X}\|^2 + \lambda \|\mathbf{X}\|_{2,1}$$

  where $\|\mathbf{X}\|_{2,1} = \sum_i \|\|\mathbf{x}_{(i)}\|$.

☹ good recovery depends on the choise of egularization parameter $\lambda > 0$

☹ not robust due to the used data fidelity term (LS-loss)

☹ not as scalable as greedy methods

Menu

Nonparametric approach based on mixed norms

$$\min_{\mathbf{X}} \underbrace{\|\mathbf{Y} - \mathbf{\Phi X}\|_{p,q}^{q}}_{\text{data fidelity}} \quad \text{subject to} \quad \underbrace{\|\mathbf{X}\|_0 \leq K}_{\text{sparsity constraint}} \qquad (P_{p,q})$$

$\Longrightarrow$ mixed ℓ_1 norms provide robustness, i.e., give larger weights on small residuals and less weight on large residuals.

We consider the cases

- $(p,q) = (1,1)$: $\|\mathbf{Y} - \mathbf{\Phi X}\|_1 = \sum_i \sum_j |y_{ij} - \boldsymbol{\phi}_{(i)}^{\mathrm{H}} \mathbf{x}_j|$

  robust norm for errors i.i.d. in space and time
- $(p,q) = (2,1)$: $\|\mathbf{Y} - \mathbf{\Phi X}\|_{2,1} = \sum_i \|\mathbf{y}_{(i)} - \mathbf{X}^{\mathrm{H}} \boldsymbol{\phi}_{(i)}\|$

  robust norm for errors i.i.d. in "space" only

and devise greedy simultaneous NIHT (SNIHT) algorithm

Normalized iterative hard thresholding (NIHT)

Conventional $\ell_{2,2}$ approach :

$$\mathbf{X}^{n+1} = \underset{\substack{\uparrow \\ \text{hard thresholding}}}{H_K} \left(\mathbf{X}^n + \overbrace{\mu^{n+1}}^{\textit{stepsize}} \underbrace{\boldsymbol{\Phi}^{\mathrm{H}}(\mathbf{Y} - \boldsymbol{\Phi}\mathbf{X}^n)}_{\substack{\uparrow \\ \text{neg. gradient } \nabla_{\mathbf{X}^*}\|\cdot\|^2}} \right)$$

- $\psi_{2,2}(\mathbf{X}) = \mathbf{X} \Longrightarrow$ Conventional $\ell_{2,2}$ approach
- $\psi_{1,1}(\mathbf{X}) = [\mathrm{sign}(x_{ij})]$ (elementwise operation of $S(\cdot)$)
- $\psi_{2,1}(\mathbf{X}) = [\mathrm{sign}(\mathbf{x}_{(i)})]$ (rowwise operation of $S(\cdot)$)

Normalized iterative hard thresholding (NIHT)

In our mixed norm $\ell_{p,q}$ approach:

$$\mathbf{X}^{n+1} = \underset{\substack{\uparrow \\ \text{hard thresholding}}}{H_K}\Big(\mathbf{X}^n + \overbrace{\mu^{n+1}}^{\text{stepsize}} \underbrace{\mathbf{\Phi}^{\mathrm{H}} \psi_{p,q}(\mathbf{Y} - \mathbf{\Phi}\mathbf{X}^n)}_{\substack{\uparrow \\ \text{neg. gradient } \nabla_{\mathbf{X}^*} \|\cdot\|_{p,q}^q}}\Big)$$

- $\psi_{2,2}(\mathbf{X}) = \mathbf{X} \Longrightarrow$ Conventional $\ell_{2,2}$ approach
- $\psi_{1,1}(\mathbf{X}) = [\mathrm{sign}(x_{ij})]$ (elementwise operation of $S(\cdot)$)
- $\psi_{2,1}(\mathbf{X}) = [\mathrm{sign}(\mathbf{x}_{(i)})]$ (rowwise operation of $S(\cdot)$)

Normalized iterative hard thresholding (NIHT)

In our mixed norm $\ell_{p,q}$ approach:

$$\mathbf{X}^{n+1} = \underset{\substack{\uparrow \\ \text{hard thresholding}}}{H_K}\Big(\mathbf{X}^n + \overbrace{\mu^{n+1}}^{\textit{stepsize}} \underbrace{\mathbf{\Phi}^{\mathrm{H}}\psi_{p,q}(\mathbf{Y} - \mathbf{\Phi}\mathbf{X}^n)}_{\substack{\uparrow \\ \text{neg. gradient } \nabla_{\mathbf{X}^*}\|\cdot\|_{p,q}^q}}\Big)$$

- $\psi_{2,2}(\mathbf{X}) = \mathbf{X} \Longrightarrow$ Conventional $\ell_{2,2}$ approach
- $\psi_{1,1}(\mathbf{X}) = [\mathrm{sign}(x_{ij})]$ (elementwise operation of $S(\cdot)$)
- $\psi_{2,1}(\mathbf{X}) = [\mathrm{sign}(\mathbf{x}_{(i)})]$ (rowwise operation of $S(\cdot)$)

Spatial sign

$$\mathrm{sign}(\mathbf{x}) = \begin{cases} \mathbf{x}/\|\mathbf{x}\|, & \text{for } \mathbf{x} \neq 0 \\ 0, & \text{for } \mathbf{x} = 0 \end{cases}$$

Normalized iterative hard thresholding (NIHT)

In our mixed norm $\ell_{p,q}$ approach:

$$\mathbf{X}^{n+1} = \underset{\substack{\uparrow \\ \text{hard thresholding}}}{H_K}\Big(\mathbf{X}^n + \overbrace{\mu^{n+1}}^{\textit{stepsize}} \underbrace{\mathbf{\Phi}^{\mathrm{H}}\psi_{p,q}(\mathbf{Y} - \mathbf{\Phi}\mathbf{X}^n)}_{\substack{\uparrow \\ \text{neg. gradient } \nabla_{\mathbf{X}^*}\|\cdot\|_{p,q}^q}}\Big)$$

- $\psi_{2,2}(\mathbf{X}) = \mathbf{X} \implies$ Conventional $\ell_{2,2}$ approach
- $\psi_{1,1}(\mathbf{X}) = [\mathrm{sign}(x_{ij})]$ (elementwise operation of $S(\cdot)$)
- $\psi_{2,1}(\mathbf{X}) = [\mathrm{sign}(\mathbf{x}_{(i)})]$ (rowwise operation of $S(\cdot)$)

Spatial sign

$$\mathrm{sign}(\mathbf{x}) = \begin{cases} \mathbf{x}/\|\mathbf{x}\|, & \text{for } \mathbf{x} \neq 0 \\ 0, & \text{for } \mathbf{x} = 0 \end{cases}$$

SNIHT(p,q) algorithm

input : $\mathbf{Y}$, $\mathbf{\Phi}$, sparsity K, mixed norm indices (p,q)
output : $(\mathbf{X}^{n+1}, \Gamma^{n+1})$ as estimates of $\mathbf{X}$ and $\Gamma = \mathrm{supp}(\mathbf{X})$
initialize: $\mathbf{X}^0 = \mathbf{0}$, $\mu^0 = 0$, $n = 0$, $\Gamma^0 = \emptyset$

$\Gamma^0 = \mathrm{supp}\big(H_K(\mathbf{\Phi}^{\mathrm{H}} \psi_{p,q}(\mathbf{Y}))\big)$
while *halting criterion false* **do**
    $\mathbf{R}_\psi^n = \psi_{p,q}(\mathbf{Y} - \mathbf{\Phi}\mathbf{X}^n)$
    $\mathbf{G}^n = \mathbf{\Phi}^{\mathrm{H}} \mathbf{R}_\psi^n$
    $\mu^{n+1} = \texttt{CompStepsize}(\Phi, \mathbf{G}^n, \Gamma^n, \mu^n, p, q)$
    $\mathbf{X}^{n+1} = H_K(\mathbf{X}^n + \mu^{n+1}\mathbf{G}^n)$
    $\Gamma^{n+1} = \mathrm{supp}(\mathbf{X}^{n+1})$
    $n = n + 1$
end

Computation of the step size μ^{n+1}

- For convergence, it is important that the stepsize is adaptively controlled at each iteration.
- Given the found support Γ^n is correct, one can choose μ^{n+1} as the min. of the objective fnc for the gradient ascent direction $\mathbf{X}^n + \mu\mathbf{G}^n|_{\Gamma^n}$:

$$L(\mu) = \left\|\mathbf{Y} - \mathbf{\Phi}\left(\mathbf{X}^n + \mu\mathbf{G}^n|_{\Gamma^n}\right)\right\|_{p,q}^q = \|\mathbf{R}^n - \mu\mathbf{B}\|_{p,q}^q$$

  where $\mathbf{R}^n = \mathbf{Y} - \mathbf{\Phi}\mathbf{X}^n$ and $\mathbf{B} = \mathbf{\Phi}\mathbf{G}_{(\Gamma^n)}$

- For $p = q$: simple linear regression problem based on ℓ_p-norm, with response $\mathbf{r} = \mathrm{vec}(\mathbf{R}^n)$ and predictor $\mathbf{b} = \mathrm{vec}(\mathbf{B})$.

$$\Rightarrow \mu^{n+1} = \|\mathbf{G}^n_{(\Gamma^n)}\|^2 / \|\mathbf{\Phi}\mathbf{G}^n_{(\Gamma^n)}\|^2 \text{ for } p = q = 2$$

- For $p = q = 1$ and $(p, q) = (2, 1)$, minimizer of $L(\mu)$ can not be found in closed-form.

Computation of the step size μ^{n+1}

- When $p = q = 1$, the solution μ verifies a fixed point (FP) equation

$$\mu = H(\mu)$$
$$H(\mu) = \Big(\sum_{i,j} |\tilde{r}_{ij}|^{-1} |b_{ij}|^2 \Big)^{-1} \sum_{i,j} |\tilde{r}_{ij}|^{-1} \mathrm{Re}(b_{ij}^* r_{ij}),$$

  and $\tilde{\mathbf{R}} = \mathbf{R}^n - \mu \mathbf{B}$ (depends on unknown μ) and $\mathbf{B} = \mathbf{\Phi} \mathbf{G}_{(\Gamma^n)}$.
- We select μ^{n+1} as the 1-step FP iterate

$$\mu^{n+1} = H(\mu^n),$$

  where μ^n is the value of the stepsize at previous nth iteration.
- Often this approximation was within $\sim 1e^{-3}$ to minimizer of $L(\mu)$.
- Same approach is used in the case $(p, q) = (2, 1)$.

Menu

Simulation study

$$\underset{M \times Q}{\mathbf{Y}} = \underset{M \times N}{\mathbf{\Phi}} \underset{N \times Q}{\mathbf{X}} + \underset{M \times 1}{\mathbf{E}}$$

Set-up

- Measurement matrix $\mathbf{\Phi}$: $\phi_{ij} \sim \mathbb{C}\mathcal{N}(0,1)$ with unit-norm columns
- $\Gamma = \mathrm{supp}(\mathbf{X})$ randomly drawn K elements from $\{1,\ldots,N\}$
- signal matrix $\mathbf{X}$: $|x_{ij}| = 10$ and $\mathrm{Arg}(x_{ij}) \sim \mathit{Unif}(0, 2\pi)$ $\forall i \in \Gamma$.

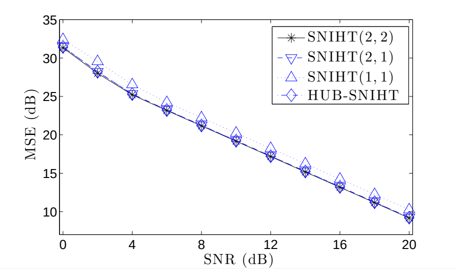
For $\ell = 1, \ldots, L$ MC-trials ($L = 1000$), compute $\mathbf{X}^{[\ell]}$ and $\Gamma^{[\ell]}$

Performance measures

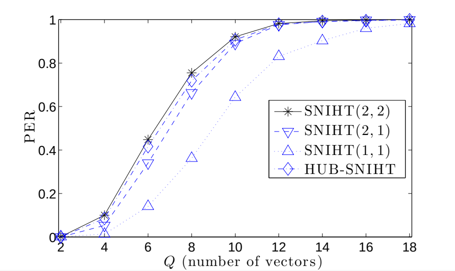
- mean squared error $\mathrm{MSE}(\hat{\mathbf{X}}) = \frac{1}{LQ} \sum_{\ell=1}^{L} \left\| \hat{\mathbf{X}}^{[\ell]} - \mathbf{X}^{[\ell]} \right\|^2$.
- probability of exact recovery $\mathrm{PER} = \dfrac{1}{L} \displaystyle\sum_{\ell=1}^{L} \mathrm{I}\big(\hat{\Gamma}^{[\ell]} = \Gamma^{[\ell]}\big)$.

IID Gaussian noise, $e_{ij} \sim \mathbb{C}\mathcal{N}(0, \sigma^2)$

MSE vs SNR for Q=16



PER vs Q at SNR = 5 dB

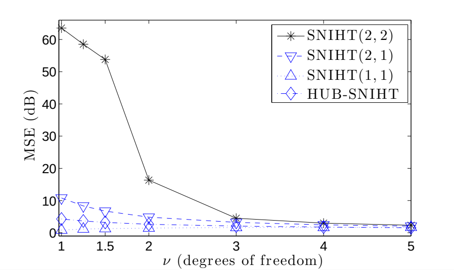


 SNIHT$(2, 2)$

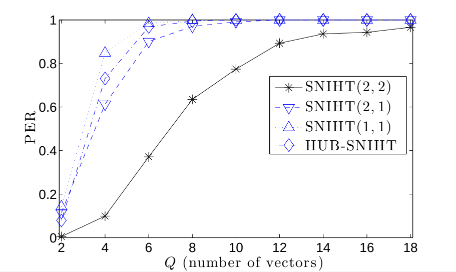
but SNIHT$(2, 1)$ and HUB-SNIHT (threshold $c = 1.517$) loose only 0.09 dB in MSE ($Q = 16$)

IID t-distributed noise, $e_{ij} \sim \mathbb{C}t_\nu(0, \sigma^2)$, $\sigma^2 = \mathrm{Med}_F(|e_i|^2)$

$\mathrm{SNR}(\sigma) = 30$ dB



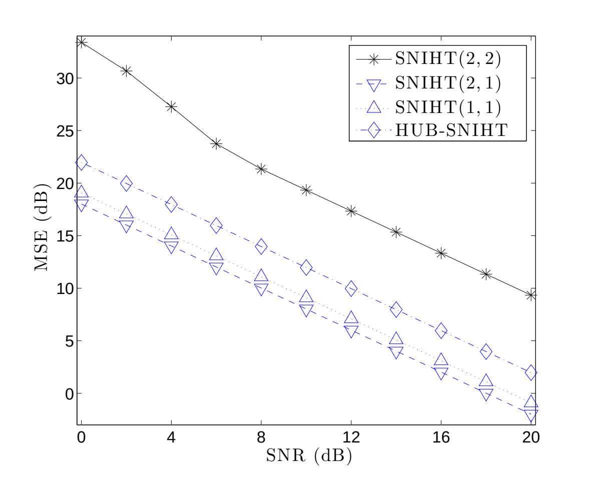
$\mathrm{SNR}(\sigma) = 10$ dB, $\nu = 3$



 SNIHT$(1,1)$

Row IIID compound Gaussian

inverse Gaussian texture, $\mathbf{e}_{(i)} \sim \mathbb{C}\mathrm{IG}_\lambda(\mathbf{0}, \sigma^2 \mathbf{I}_Q)$



SNIHT(2, 1)

Menu

Sensor array and narrowband, farfield source model

- M sensors, K sources ($M > K$):
- Model for sensor measurements at time instant t

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(t) + \mathbf{e}(t),$$

- steering matrix $\mathbf{A}(\boldsymbol{\theta}) = \begin{pmatrix}\mathbf{a}(\theta_1) & \ldots & \mathbf{a}(\theta_K)\end{pmatrix}$
- unknown (distinct) Direction of Arrivals (DOA's) $\theta_1, \ldots, \theta_K$
- known array manifold $\mathbf{a}(\theta)$ (e.g., ULA)
- unknown signal waveforms $\mathbf{x}(t) = (x_1(t), \ldots, x_K(t))^\top$

Objective

Estimate the DOA's of the sources given the snapshots $\mathbf{y}(t_1), \ldots, \mathbf{y}(t_Q)$ and the number of sources K impinging on the sensor array.

Multichannel sparse recovery approach to source localization

- Construct an $M \times N$ overcomplete steering matrix $\mathbf{A}(\tilde{\theta})$ for a grid of all source locations $\tilde{\theta}_1, \ldots, \tilde{\theta}_N$ of interest.
- Suppose that $\tilde{\theta}$ contains the true DOA's to some accuracy.
- The array model in matrix form then rewrites as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}(t_1) & \cdots & \mathbf{y}(t_Q) \end{pmatrix} = \mathbf{A}(\tilde{\theta})\mathbf{X} + \mathbf{E},$$

where $\mathbf{X} \in \mathbb{C}^{N \times Q}$ is K-sparse, with K non-zero rows containing the source waveforms at time instants $t_1, \ldots, t_Q$.

$\Longrightarrow$ K-sparse MMV model

- In the multichannel sparce recovery formulation $\mathbf{A}(\tilde{\theta})$ is completely known, and we can use SNIHT methods to identify the support.

Multichannel sparse recovery approach to source localization

- Construct an $M \times N$ overcomplete steering matrix $\mathbf{A}(\tilde{\theta})$ for a grid of all source locations $\tilde{\theta}_1, \ldots, \tilde{\theta}_N$ of interest.
- Suppose that $\tilde{\theta}$ contains the true DOA's to some accuracy.
- The array model in matrix form then rewrites as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}(t_1) & \cdots & \mathbf{y}(t_Q) \end{pmatrix} = \mathbf{A}(\tilde{\theta})\mathbf{X} + \mathbf{E},$$

where $\mathbf{X} \in \mathbb{C}^{N \times Q}$ is K-sparse, with K non-zero rows containing the source waveforms at time instants $t_1, \ldots, t_Q$.

$\Longrightarrow$ K-sparse MMV model

- In the multichannel sparce recovery formulation $\mathbf{A}(\tilde{\theta})$ is completely known, and we can use SNIHT methods to identify the support.

Multichannel sparse recovery approach to source localization

- Construct an $M \times N$ overcomplete steering matrix $\mathbf{A}(\tilde{\boldsymbol{\theta}})$ for a grid of all source locations $\tilde{\theta}_1, \ldots, \tilde{\theta}_N$ of interest.
- Suppose that $\tilde{\boldsymbol{\theta}}$ contains the true DOA's to some accuracy.
- The array model in matrix form then rewrites as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}(t_1) & \cdots & \mathbf{y}(t_Q) \end{pmatrix} = \mathbf{A}(\tilde{\boldsymbol{\theta}})\mathbf{X} + \mathbf{E},$$

where $\mathbf{X} \in \mathbb{C}^{N \times Q}$ is K-sparse, with K non-zero rows containing the source waveforms at time instants $t_1, \ldots, t_Q$.

$\Longrightarrow$ Finding DOA's is equivalent to identifying $\Gamma = \mathrm{supp}(\mathbf{X})$

- In the multichannel sparse recovery formulation $\mathbf{A}(\tilde{\boldsymbol{\theta}})$ is completely known, and we can use SNIHT methods to identify the support.

Multichannel sparse recovery approach to source localization

- Construct an $M \times N$ overcomplete steering matrix $\mathbf{A}(\tilde{\boldsymbol{\theta}})$ for a grid of all source locations $\tilde{\theta}_1, \ldots, \tilde{\theta}_N$ of interest.
- Suppose that $\tilde{\boldsymbol{\theta}}$ contains the true DOA's to some accuracy.
- The array model in matrix form then rewrites as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}(t_1) & \cdots & \mathbf{y}(t_Q) \end{pmatrix} = \mathbf{A}(\tilde{\boldsymbol{\theta}})\mathbf{X} + \mathbf{E},$$

  where $\mathbf{X} \in \mathbb{C}^{N \times Q}$ is K-sparse, with K non-zero rows containing the source waveforms at time instants $t_1, \ldots, t_Q$.

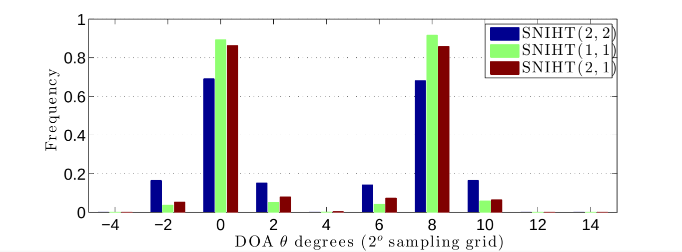
  $\Longrightarrow$ Finding DOA's is equivalent to identifying $\Gamma = \mathrm{supp}(\mathbf{X})$
- In the multichannel sparse recovery formulation $\mathbf{A}(\tilde{\boldsymbol{\theta}})$ is completely known, and we can use SNIHT methods to identify the support.

Simul set-up

- ULA of $M = 20$ sensors with half a wavelength interelement spacing.
- $K = 2$ independent Gaussian sources from $\theta_1 = 0^o$ and $\theta_2 = 8^o$.
- Temporally/spatially white Gaussian error terms.
- Low SNR $= 0$ dB.
- Uniform grid $\tilde{\theta}$ on $[-90, 90]$ with 2^o degree spacing
- We compute PER rates and relative frequencies of found DOA estimates in the grid for 1000 Monte Carlo runs.

Number of snapshots $Q = 4$

DOA's at $\theta_1 = 0^o$ and $\theta_1 = 8^o$.



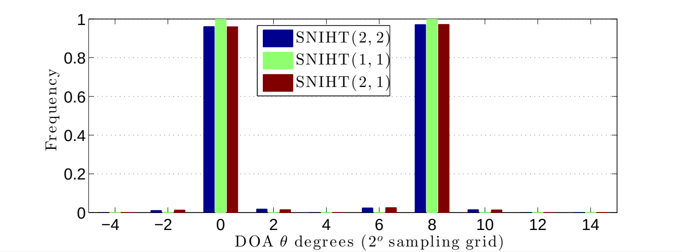
<u>PER rates</u>

1. $\text{SNIHT}(1,1) = 81\%$
2. $\text{SNIHT}(2,1) = 73.1\%$
3. $\text{HUB-SNIHT} = 42.6\%$
4. $\text{SNIHT} = 37.6\%$

$\text{SNIHT}(1,1)$

Number of snapshots $Q = 40$

DOA's at $\theta_1 = 0^o$ and $\theta_1 = 8^o$.

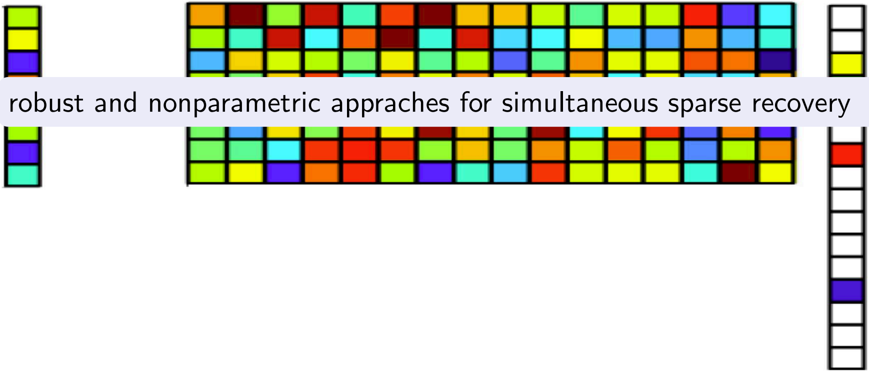


PER rates

1. $\text{SNIHT}(1,1) = 100\%$
2. $\text{SNIHT}(2,1) = 73.1\%$
3. HUB-SNIHT $= 42.6\%$
4. SNIHT $= 37.6\%$

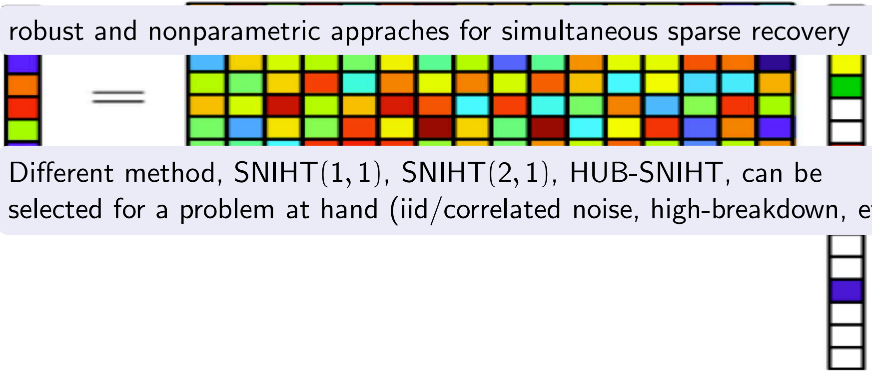
$\text{SNIHT}(1,1)$

Conclusions



robust and nonparametric appraches for simultaneous sparse recovery

Conclusions

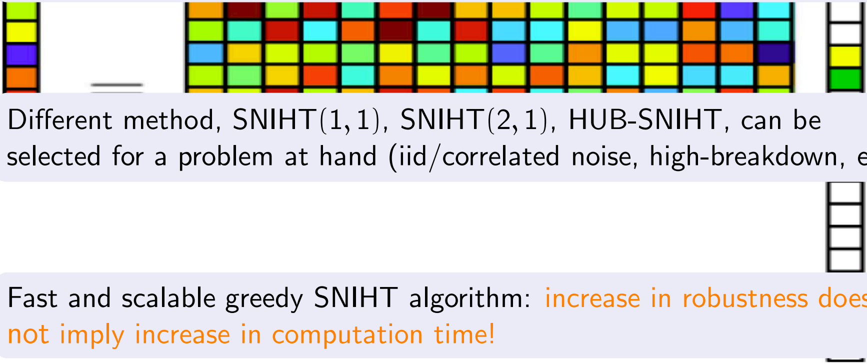


robust and nonparametric appraches for simultaneous sparse recovery

Different method, SNIHT$(1,1)$, SNIHT$(2,1)$, HUB-SNIHT, can be selected for a problem at hand (iid/correlated noise, high-breakdown, etc)

Conclusions

robust and nonparametric appraches for simultaneous sparse recovery

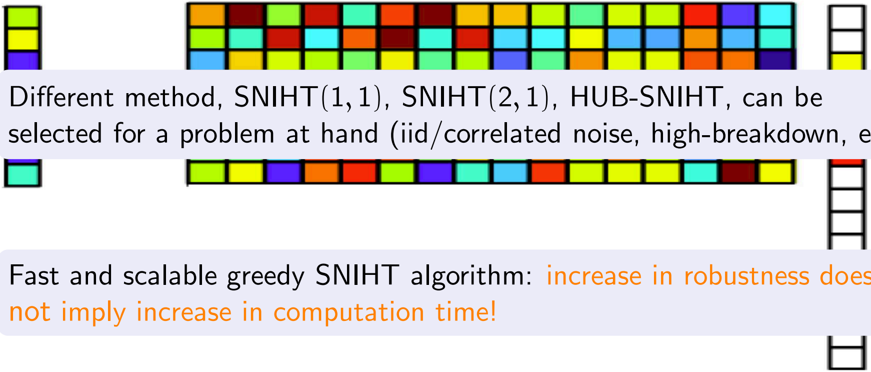


Different method, SNIHT$(1, 1)$, SNIHT$(2, 1)$, HUB-SNIHT, can be selected for a problem at hand (iid/correlated noise, high-breakdown, etc)

Fast and scalable greedy SNIHT algorithm: increase in robustness does not imply increase in computation time!

Conclusions

robust and nonparametric appraches for simultaneous sparse recovery



Different method, SNIHT$(1,1)$, SNIHT$(2,1)$, HUB-SNIHT, can be selected for a problem at hand (iid/correlated noise, high-breakdown, etc)

Fast and scalable greedy SNIHT algorithm: increase in robustness does not imply increase in computation time!

Applications in source localization and image denoising, etc.

This talk based on

E. Ollila (2015a)
Nonparametric Simultaneous Sparse Recovery: an Application to Source Localization
PDF (ArXiv): http://arxiv.org/pdf/1502.02441v1
EUSIPCO'15, submitted.

E. Ollila (2015b).
Robust Simultaneous Sparse Approximation
A Festschrift in Honor of Hannu Oja, Springer, Sept. 2015.

Blanchard, J. D., Cermak, M., Hanle, D., and Jin, Y. (2014).
Greedy algorithms for joint sparse recovery.
IEEE Trans. Signal Processing, 62(7).

Chen, J. and Huo, X. (2006).
Theoretical results on sparse representations of multiple-measurement vectors.
IEEE Trans. Signal Processing, 54(12):4634–4643.

Cotter, S. and Rao, B. (2002).
Sparse channel estimation via matching pursuit with application to equalization.
IEEE Trans. Comm., 50(3):374–377.

Duarte, M. F. and Eldar, Y. C. (2011).
Structured compressed sensing: From theory to applications.
IEEE Trans. Signal Processing, 59(9):4053–4085.

Eldar, Y. C. and Rauhut, H. (2010).
Average case analysis of multichannel sparse recovery using convex relaxation.
IEEE Trans. Inform. Theory, 56(1):505–519.

Gorodnitsky, I., George, J., and Rao, B. (1995).
Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm.
J. Electroencephalogr. Clin. Neurophysiol., 95(4):231–251.

Gribonval, R. and Zibulevsky, M. (2010).
Handbook of Blind Source Separation, chapter Sparse component analysis, pages 367–420.
Academic Press, Oxford, UK.

Malioutov, D., Çetin, M., and Willsky, A. S. (2005).
A sparse signal reconstruction perspective for source localization with sensor arrays.
IEEE Trans. Signal Processing, 53(8):3010–3022.

Tropp, J. A. (2006).
Algorithms for simultaneous sparse approximation. part II: Convex relaxation.
Signal Processing, 86:589–602.